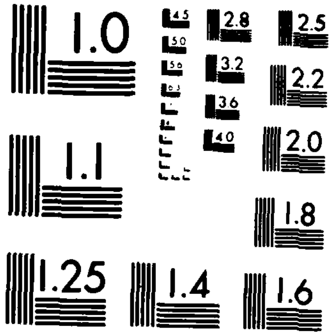
AD-A193 627 CONJUNCTIVE MEASUREMENT THEORY: COGNITIVE RESEARCH PROSPECTS(U) SOUTH CAROLINA UNIV COLUMBIA CENTER FOR MACHINE INTELLIGENCE R J JANNARONE 25 JAN 88 1/1

UNCLASSIFIED USCMI-88-12 N00014-86-K-0817 F/G 5/8 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Conjunctive Measurement Theory: Cognitive Research Prospects†

Robert J. Jannarone

University of South Carolina

USCMI Report No. 88-12

CENTER FOR MACHINE INTELLIGENCE

UNIVERSITY OF SOUTH CAROLINA

COLUMBIA, SC 29208

This research was sponsored by Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-86-K00817, Authority Identificati[on] Number, NR 4421-544.

4

Conjunctive Measurement Theory:

Cognitive Research Prospects†

Robert J. Jannarone

University of South Carolina

USCMI Report No. 88-12

25 January, 1988

Key words: Item response theory; conjunctive models, compensatory, reactive measurement, nonadditive measurement, Rasch model.

† Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government. This research was sponsored by Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research Contract No. N00014-86-K00817, Authority Identification Number, NR 4421-544.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) ONR 86-2	5. MONITORING ORGANIZATION REPORT NUMBER(S) Office of Naval Research
6a NAME OF PERFORMING ORGANIZATION University of South Carolina; 6b OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION 800 N. Quincy St., Code 442 Arlington, VA 22217
6c. ADDRESS (City, State, and ZIP Code) Columbia, South Carolina 29208	7b. ADDRESS (City, State, and ZIP Code)
8a NAME OF FUNDING/SPONSORING ORGANIZATION Personnel & Training Res.; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K00817
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61153N 42	RR04204	RR0420401	4421-544

11. TITLE (Include Security Classification)
Conjunctive Measurement Theory: Cognitive Research Prospects

12. PERSONAL AUTHOR(S)
Robert J. Jannarone

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 8-15-86 TO 11-31-87		30

16 SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

A new psychometric development theory based on conjunctive measurement is introduced with an eye toward cognitive modeling. Conjunctive measures are shown that can reflect persons' abilities to: combine component skills that are individually necessary for solving a given task; learn information at one point and successfuly use it at a later point; positively transfer learned material from one setting to another; and improve knowledge in different ways between pretests and posttests. Several theoretical issues are also introduced to contrast conjunctive testing with traditional testing. The main distinction is that conjunctive measurement violates test theory's local independence axiom, by allowing individuals to change rather than be unaffected during the measurement process.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Charles E. Davis	22b TELEPHONE (Include Area Code) (202) 696-4046; 22c OFFICE SYMBOL

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Table of Contents

Introduction

Scope. Cognitive theories have shown an increasing complexity trend throughout their history. Notable examples are the early trend from single to multiple ability traits (Spearman, 1904; Thurstone, 1938; Guilford, 1967), the analysis of analogical reasoning tasks into component subtasks (Sternberg, 1977; Embretson, 1984); and recent new advances in test design (Embretson, 1985). In reaction to this trend in ***cognitive*** theory, ***psychometric*** theorists have proposed increasingly detailed test models. Notable examples include the advent of multiple factor analysis (Thurstone, 1932; Jöreskog, 1978); and multivariate item response theory (Fischer, 1973; Andersen, 1980;, Embretson, 1985). Cognitive and psychometric theories have thus shown increasing trends in both generality and mutual alignment.

The recent development of conjunctive item response theory (Jannarone, 1986; 1988a; 1988b; Jannarone, Laughlin, & Yu, 1988) suggests new areas for psychometric model expansion. The conjunctive modeling development has partly been in reaction to the componential analogical reasoning movement—both involve component abilities that are individually necessary for solving a composite task, hence ***conjunctively*** related. Conjunctive measurement theory has not yet influenced cognitive modeling, however, because it is rather new and has so far only been presented in mathematical form. Yet the structures that conjunctive measurement reflects are both closely related to modern cognitive work and clearly distinct from more traditional psychometric structures. Some interesting related prospects for cognitive research thus seem possible.

This report is an attempt to describe conjunctive test theory, to contrast it with other test theories, and to encourage related psychometric and cognitive future developments.

As will be shown with a variety of examples, conjunctive measurement has the potential for: uncovering conjunctive cognitive structures; measuring different problem solving styles; measuring person's abilities to learn information at one point and successfully apply it at a later point; evaluating person's uses of alternative learning styles; and providing realistic models for computer aided instruction settings.

Several distinctions between conjunctive and traditional test items will also be described. The most theoretically important distinction is that conjunctive measurement allows persons to change as a part of the measurement process. This permits traits such as learning styles to be measured, but it also marks a basic departure from traditional test theory's axioms.

Purpose. One goal of this report is to describe the major distinctions between conjunctive and traditional measurement theories. A second goal is to show how conjunctive measurement can be useful, by describing its key features within ability assessment settings.

In the following sections I will first give some examples of conjunctive ability settings, structural models, and procedural guidelines. I will follow with some theoretical perspectives and finish with some future directions for psychometric and cognitive research.

Conjunctive Measurement Overview

Some examples. I will give three examples next—one involving two component abilities that have conjunctive effects on a composite ability; one based on a chain of items that are linked by sequential learning effects; and one involving replicated tests that have conjunctively linked pretest and posttest items. For now, the common elements to look for in these examples are that: (a) each involves component items that are linked together by underlying cognitive tasks. (b) each measures individual differences in item linkages; and (c) each leads to measures of item linkages that are nonadditive functions of item cross-product scores, rather than additive functions of item scores.

The first example involves measuring component abilities and evaluating their joint effects on analogical reasoning. Table 1 contains three items that are designed to reflect analogical reasoning abilities (kindly provided by Susan Embretson—see Embretson (Whitely), 1984). Such items are presented to subjects in triplets like that in Table 1. The Total item represents overall analogical

Table 1

Information component subtasks for verbal analogies

Total Item	Fist:Clench::Teeth:______________ 1) Pull 2) Brush 3) Grit 4) Gnaw 5) Jaw
Rule Construction Subtask	Fist:Clench::Teeth:______________ Rule:______________
Response Evaluation Subtask	Fist:Clench::Teeth:______________ Rule: Angry reaction done with "teeth" 1) Pull 2) Brush 3) Grit 4) Gnaw 5) Jaw Circle answer that best fulfills the given rule

reasoning ability, whereas the Rule Construction and the Response Evaluation items represent two component subtest abilities. Tests made up of such item triplets have been studied in the past (ibid: Pellegrino & Glaser, 1979; Pellegrino, Mumaw, & Shute, 1985; Sternberg, 1977) to show how subtask skills are used in solving analogies. I will focus on how persons' responses might indicate whether (a) both subtask skills are necessary for passing a Total item; or (b) only one of the subtask skills may be sufficient for passing the Total item.

Suppose that scores were available from a group of persons who were tested on N such item triplets. Traditional test construction methods would suggest that three subscales be formed, each being based on N out of the $3N$ items. In their simplest form the subscales would combine their item scores additively and equally, yielding,

$$s^{(C)} = \sum_{n=1}^{N} x_n^{(C)},\ s^{(E)} = \sum_{n=1}^{N} x_n^{(E)},\ s^{(T)} = \sum_{n=1}^{N} x_n^{(T)}, \tag{1}$$

where the three sums indicate the number of correct Rule Construction, Response Evaluation, and Total items, respectively. (The items are meant to be coded in the usual binary way, *i.e.* $x_n^{(C)}$, $x_n^{(E)}$, $x_n^{(T)} = 0$ for PASS, 1 for FAIL.)

The traditional additive scoring formulas shown in (1) could be useful, up to a point. The relative additive impacts of each subtask ability on total ability could be evaluated separately as well as stepwise. The effects of the three subscales on external criteria could also be assessed by using standard factor analysis and regression methods.

Some interesting response pattern differences could not be reflected by additive subscales, however. For example two viable strategies could exist for passing a Total item. One strategy might require that both the Rule Construction skill and the Response Evaluation skill be available for passing each Total item. However the other strategy might require only one of the subtask skills, perhaps along with other unmeasured skills. Suppose that 18 item triplets were presented to a group of people and that a subgroup responded correctly to exactly 6 items of each type. Thus, each person in the subgroup would earn number-correct scores of 6, 6, and 6 out of 18 $s^{(C)}$, $s^{(E)}$, and $s^{(T)}$ items, respectively. Any analysis based only on those scores alone could not distinguish the responses among any persons in the subgroup. Yet, different subsample members might use different strategies consistently. In the extreme, the scores on each item triplet for some persons would be 1 if and only if their Total item score on the triplet were 1. Such response patterns would clearly indicate the use of a strategy that required *both* subtask skills. For other persons, passing Total items would always coincide with passing only one of the two subtask items, indicating another strategy.

These kinds of distinct strategies could not be reflected by additive scales, but rather by nonadditive subscales of the form,

$$s^{(CE)} = \sum_{n=1}^{N} x_n^{(C)} x_n^{(E)} . \tag{2}$$

For example in the subsample of persons who had additive scores of 6, 6, and 6 on the 18-item test, those requiring both subtasks would have $s^{(CE)}$ values of 6. By contrast, those requiring only one subskill would have lower $s^{(CE)}$ values. Formal logic can also be used to contrast different types of measurement in such cases. In logical terms distinct strategies would be reflected by distinct *conjuncts* among the component item events (PASS = TRUE, FAIL = FALSE). For example, those having many TRUE values for three-way conjuncts among item triplets would reflect one strategy; whereas those having few TRUE values would reflect another. This is my basis for referring to the models based on (2) and on similar measures as *conjunctive*.

More complex subscales could be used that were based on all possible conjuncts among the three subscale items, for example by going beyond item triplet boundaries. However, these would be difficult to deal with, both statistically and conceptually. Similar concerns hold for the two examples to follow.

The second example is a test made up of items that are linked together into a chain by adjacent interitem dependencies. The first three items in the chain are given in Table 2. (The key words for these items were kindly suggested by Chris McCormick and Gloria Miller—see McCormick and Miller,

Table 2. Three Possible Linked Learning Items

1. PADLE

For gardening, the most common earth-moving tasks are digging, smoothing, breaking clods, and furrowing. Therefore, a gardener's tools should include a shovel, a rake and a padle.

What is a padle?

(a) a spade

(b) a pickaxe

(c) a hoe

(d) a mower

2. PADLE ∩ KAVA

The term 'root beer' may be misleading, unless the beverage happens to be made from kava.

How can a padle be instrumental to having a good time?

(a) through distilling kava

(b) through harvesting kava

(c) through transporting kava

(d) through weaving kava

3. KAVA ∩ CANGUE

In medieval Asia, using a cangue on a prisoner would often result in a quick confession, unless perhaps the guards had provided him with kava.

How could the kava intervene?

(a) by poisoning the prisoner

(b) by arming the prisoner

(c) by intoxicating the prisoner

(d) by befriending the prisoner

4. CANGUE ∩ ____

•
•
•

1986). Item 1 tests for comprehension in the usual way by first introducing a word (PADLE) and then testing whether or not the word's meaning was correctly learned. Item 2 is unusual, however, because passing it requires that both among two words be learned—one word (KAVA) that is introduced in Item 2, but another word (PADLE) that is introduced in Item 1. Likewise, Item 3 tests whether or not both the word introduced in Item 2 and the word introduced in Item 3 are learned. Other items in the test similarly evaluate whether both the word from an item and the word from the immediately preceding item are learned. The resulting structure of such a test is a chain made up of adjacent items that are linked together semanticly.

The dependencies among items for such a test link adjacent items so that persons' abilities to ***effectively learn*** may be measured. By effective learning I mean learning something new as well as successfully applying it later. Measuring effective learning ability would be potentially useful in selecting training programs, studying learning skills, and diagnosing learning impairments.

The most direct way to solve the items in Table 2 would be to learn both the meaning of the new concept for one item and the concept from the preceding item. The most direct way to measure this item-solving style, in turn, would be to evaluate the following ***adjacent cross-product*** score for an M-item test:

$$s^{(d)} = \sum_{m=1}^{M-1} x_m x_{m+1} \ . \tag{3}$$

As in the previous example, alternative item passing styles might also be possible. For example some people might tend to learn the correct word meaning for an item only after having thought about its usage in the next item. In that case more items might be passed than the value of d in (3) might indicate. Also as in the previous example, useful information could be obtained by only using the additive alternatives to (3), such as persons' usual number-correct scores,

$$s^{(g)} = \sum_{m=1}^{M} x_m \ . \tag{4}$$

For example, additive scores would provide the best single measures of overall test performance.

A key issue for conjunctive modeling is the extent that nonadditive scoring adds information to traditional additive scoring. Table 3 illustrates the extra potential for nonadditive information in the item chain learning case. The row margins in Table 3 give the number of test patterns that could lead to number-correct scores from 0 to 15 on a 15-item test. (They also give the expected number of persons out of 32,768 who would get different g values if all such patterns were equally likely.) Each row in Table 3 breaks down its g contingency into possible $(s^{(g)}, s^{(d)})$ contingencies. For example, if g were 6 then d could have possible values between 0 and 5, as indicated by the corresponding row in the table.

The potential for added information shown in Table 3 is similar to that for cross-product scores from the previous analogy example. Suppose that possible associations were of interest between some external measure and performance on a test having this kind of item chain structure. An analysis based on g alone from a 15-item test would allow 16 groups of people to be compared on the external measure. Including d in the analysis as well, however, could lead to a much finer breakdown. For example among the subgroup having g values of 6, five smaller subgroups could be compared on the external criterion, and so on for the other possible g values. Using d along with g could also be substantively interesting insofar as different d values might reflect distinct strategies and skills.

Tests having serially dependent item structures can reflect other traits that traditional models cannot. These include: (a) settings where some persons may have positive learning transfer (*e.g.* learning on one item that improves the likelihood of passing later items) but others may have negative learning transfer; (b) cases where students perform worse after some training than they did at the outset—for example when people knew inferior techniques prior to training; and (c) cases where clearly brighter persons perform worse than less bright persons, because they "think themselves into a jam"—for example from being distracted by some incorrect item choices. The power of conjunctive models for reflecting such traits is illustrated in Figure 1. The figure contains passing probabilities as functions of ability for one item from a test that follows a certain item chain structure (Jannarone.

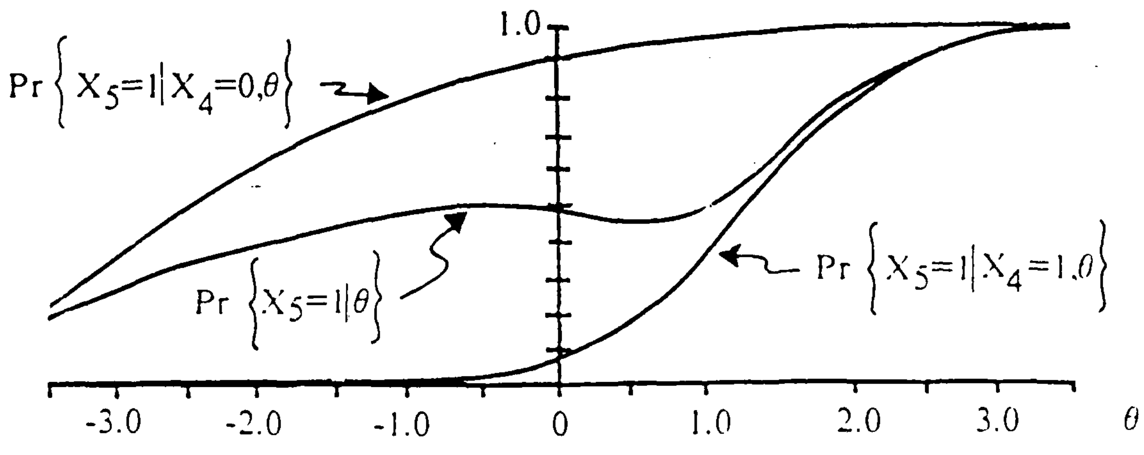
Table 3.

Contingencies among Test score patterns yielding Distinct Joint and Marginal g, d values (M=15)*

		d														
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Marginal g Frequencies
	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	1	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15
	2	91	14	0	0	0	0	0	0	0	0	0	0	0	0	0	105
	3	286	156	13	0	0	0	0	0	0	0	0	0	0	0	0	455
	4	495	660	198	12	0	0	0	0	0	0	0	0	0	0	0	1365
	5	462	1320	990	220	11	0	0	0	0	0	0	0	0	0	0	3003
	6	210	1260	2100	1200	225	10	0	0	0	0	0	0	0	0	0	5005
	7	36	504	1890	2520	1260	216	9	0	0	0	0	0	0	0	0	6435
g	8	1	56	588	1960	2450	1176	196	8	0	0	0	0	0	0	0	6435
	9	0	0	28	392	1470	1960	980	168	7	0	0	0	0	0	0	5005
	10	0	0	0	0	126	756	1260	720	135	6	0	0	0	0	0	3003
	11	0	0	0	0	0	0	210	600	450	100	5	0	0	0	0	1365
	12	0	0	0	0	0	0	0	0	165	220	66	4	0	0	0	455
	13	0	0	0	0	0	0	0	0	0	0	66	36	3	0	0	105
	14	0	0	0	0	0	0	0	0	0	0	0	0	13	2	0	15
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Marginal d Frequencies		1597	3970	5807	6304	5542	4118	2655	1496	757	326	137	40	16	2	1	Total 32,768

* Each entry is the number of distinct test patterns from a 15-item test that could yield the indicated joint and marginal g and d values.

Figure 1. Response Probabilities For One Item, Given a Particular Univariate Rasch Markov Model.



1988a). Three probability functions for the items are shown: one that is conditional on the previous item having been failed, one that is conditional on the previous item having been passed, and the third that is unconditional. The two conditional graphs show that for all ability levels the probability of passing item 5 is always higher if item 4 was passed than if it was failed. The two conditional graphs thus indicate positive transfer for all persons, reflecting a trait like that suggested in (a) above. The unconditional graph indicates that along part of the ability range the probability of passing the item *goes down* as ability increases. It thus permits traits to impact upon items as in (b) and (c).

This conjunctive modeling potential for reflecting traits like (a) through (c) is notable because no traditional test models have the same potential. By contrast traditional models require performance to be an increasing function of ability, thus ruling out settings like (b) and (c). They also require that persons' item passing probabilities be independent of their other item scores. This rules out the possibility for reflecting learning transfer effects as in (a).

The third example concerns settings where the same test or test battery is given on different occasions. Figure 2 shows some possible responses from a battery of ten subtests taken on two different dates (pretest and posttest). Each of the five graphs in Figure 2 is a possible scatterplot ***for a given person***, with that person's ten pretest, posttest scores each marked by an $\times$ in the graph. The five graphs are similar in that they share the same ten pretest scores. Also, all five of the graphs show the same average improvement of the ten posttest scores over the ten pretest scores. The five graphs differ, however, in the ways that the pretest and posttest scores are correlated. The Figure 2c graph shows a person whose posttest scores and pretest scores are uncorrelated. The other four graphs show persons having pretest and posttest scores that are correlated, but in different ways.

The graphs in Figure 2 might indicate different strategies that people might use. For example suppose that five persons were given a diagnostic test battery and then allowed to study before retaking the test battery. Figure 2a shows a person who would decide to improve on each topic uniformly; Figure 2b shows a person who would choose to maximize his or her minimum posttest score; Figure 2d shows a person who would decide to excel on her/his best pretest scores; and Figure 2e shows a person who would choose to excel on her/his worst pretest scores.

As in the previous examples, additive measurement can be useful in the pretest-posttest case, up to a point. The most popular way to analyze such scores is to construct additive pretest scores and additive posttest scores of the form,

$$s^{(pre)} = \sum_{k=1}^{K} x_k^{(pre)} , \quad s^{(post)} = \sum_{k=1}^{K} x_k^{(post)} \tag{5}$$

Given such scores, individual differences can be explored in pretest scores, posttest scores, and change scores of the from, $s^{(post)} - s^{(pre)}$. Moreover, individual differences in such change scores can be quite informative (Rogosa & Willett, 1982). However, the additive statistics in (5) may not reflect some interesting data features. For example, all of the five graphs in Figure 2 have been constructed to each have the same pretest and posttest scores, hence the same change scores. Yet the different patterns among the graphs point toward some distinct strategies and styles, as stated earlier.

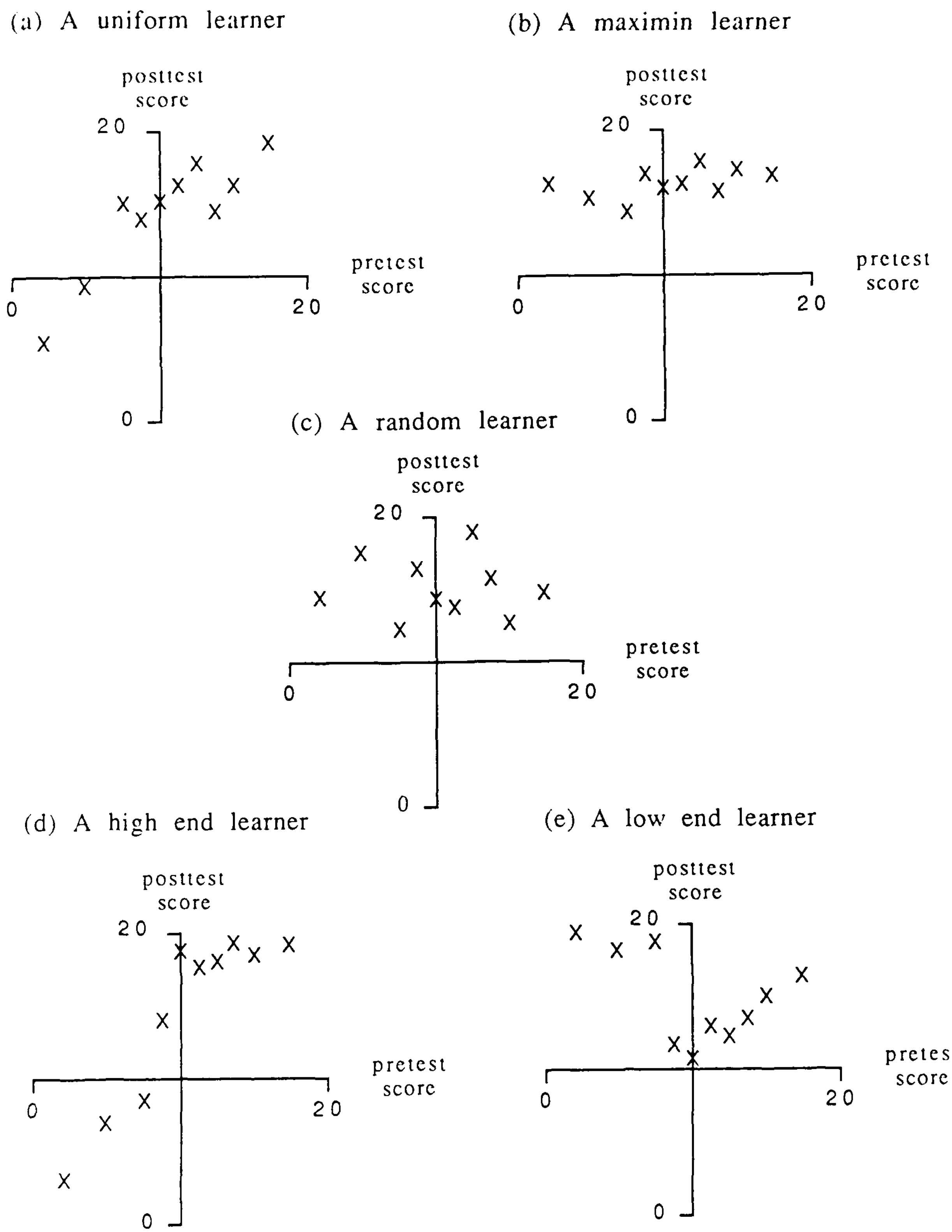
One way to capture the distinctions among the Figure 2 graphs is to evaluate nonadditive statistics of the form,

$$s^{(pre,post)} = \sum_{k=1}^{K} x_k^{(pre)} x_k^{(post)} . \tag{6}$$

For example, Person correlation coefficients based on the cross-product statistic in (6) could be computed among the ten pretest and posttest items for each graph. The correlation coefficients would have distinct values for the five graphs, thus providing a means for reflecting the five different styles.

The Figure 2 graphs show how standard pretest posttest formats may be used for measuring something novel – a kind of learning style. The graphs also reflect an interesting psychometric property. They all violate what is often called the fundamental axiom of test theory—the (local independence) requirement that for a given person no item subtest measures can depend on any others. This important distinction will be discussed later.

Figure 2. Some Hypothetical Learning Styles



These examples have been presented with an emphasis on three distinct features: (a) tests can be formed within standard binary item formats, yet with substantively interesting item response dependencies; (b) individual differences in these dependencies may be of interest; and (c) measuring such individual differences requires nonadditive rather than traditional additive scoring formulas. All three examples focus on different strategies that persons might use, which are based on and reflected by conjunctive item information. Indeed, this potential for reflecting how persons can *react to*—rather than simply be measured by—items separates conjunctive from additive measurement.

Conjunctive ability structure. (This section contains mathematical details that some readers may wish to skip.) I will begin by describing the structure for a general class of conjunctive models, which includes all three of the above examples as special cases. The general form involves M component subtests or items—for the analogy example, $M = 3N$; for the pretest-posttest example, $M = 2K$; and M for the chain linked learning case is the same as M for the general case. For all cases the likelihood that a given person will have a given item response pattern is an exponential function having an exponent of the form,

$$\sum_{(m_1, \ldots, m_s) \in \tilde{G}} (\theta^{(m_1, \ldots, m_s)} - \beta_{m_1, \ldots, m_s}) x_{m_1}, \ldots, x_{m_s} , \tag{7}$$

where the $\theta^{(m_1, \ldots, m_s)}$ are person parameters, the $\beta_{m_1, \ldots, m_s}$ are item parameters, the x's are item scores, and $\tilde{G} = \{1, 2, \ldots, M, (1,2), (1,3), \ldots, (1, \ldots, M)\}$.

The terms in (6) involve M terms that are first-order functions of the items,

$$\binom{M}{2}$$

terms that are second-order functions of the items, and so on. As mentioned earlier such terms represent logical conjuncts among item performance events, because the items are binary. The first-order terms in (7) are thus related to first-order conjuncts, the second-order terms to second-order conjuncts, and so on.

The total possible number of such terms in the exponent of (7) is 2^M, a prohibitively large number unless M is small. Consequently, for most models of interest (including all examples in this report) the weights for most of the terms are set to 0.

Another way of specifying the general form is to express (7) as

$$\sum_{(m_1, \ldots, m_s) \in \tilde{R}} (\theta^{(m_1, \ldots, m_s)} - \beta_{m_1, \ldots, m_s}) x_{m_1} \cdots x_{m_s} , \tag{8}$$

where $\tilde{R} \subseteq \tilde{G}$. The family of models satisfying (8) are called the conjunctive Rasch family (*CRF*), for reasons that will become clear later. Since each distinct exponent defines a distinct model, each special case of (8) is called its corresponding model's *label*.

Before describing the previous examples as special cases of the *CRF*, I need to introduce one additional simplifying device. As stated (8) includes too many individual parameters to be practical. All useful versions of the *CRF* reduce individual parameters to manageable numbers, by fixing some parameter values at zero and/or forcing some to always equal others. For example the well known (additive) Rasch item response model is a special case of the *CRF*. The label for the Rasch model takes the form,

$$\sum_{m=1}^{M} (\theta - \beta_m) x_m . \tag{9}$$

The Rasch model (9) can be recognized as a special case of (8) by noticing that (a) $\tilde{R}$ for the Rasch model case is simply $\{1, 2, \ldots, M\}$; and (b) the individual parameters are reduced to only one common parameter by setting $\theta^{(1)} = \theta^{(2)} = \cdots = \theta^{(M)} = \theta$. The remaining examples are also such special cases of (8).

For all special cases of the *CRF* conjunct probabilities are increasing functions of their person parameters and decreasing functions of their item parameters. In the Rasch case each first-order

conjunct (item score) has a probability of being TRUE (PASS) that is an increasing function of its person parameter (ability) and a decreasing function of its item parameter (difficulty). Probabilities for higher-order conjuncts in the examples to follow behave similarly, with the higher-order conjunct values indicating that each component item was passed.

Given the existence of random samples based on I individuals, each version of (8) leads to a corresponding ***sample likelihood.*** (Each person's parameter value and item score in the sequel will be denoted by an I subscript, with I ranging from 1 to I.) Each sample likelihood, in turn, includes a set of ***sufficient statistics*** for the model. A model's sufficient statistics are the only statistics that need be computed from the raw item scores in order to analyze the model statistically. Fortunately, sufficient statistics based on conjunctive models satisfying (8) are very easy to compute and interpret, as will be shown next.

Beginning with the analogies example, the additive version would have the label,

$$\sum_{n=1}^{N} (\theta^{(C)}-\beta_n^{(C)})x_n^{(C)} + \sum_{n=1}^{N} (\theta^{(E)}-\beta_n^{(E)})x_n^{(E)} + \sum_{n=1}^{N} (\theta^{(T)}-\beta_n^{(T)})x_n^{(T)} . \tag{10}$$

Person sufficient statistics based on additive analogies subscales would be the usual number-correct subscale scores,

$$s_i^{(C)} = \sum_{n=1}^{N} x_{in}^{(C)},\ s_i^{(E)} = \sum_{i=1}^{N} x_{in}^{(E)},\ s_i^{(T)} = \sum_{i=1}^{N} x_{in}^{(T)},\ i = 1, \ldots, I , \tag{11}$$

and additive item difficulty statistics would be,

$$t_n^{(C)} = \sum_{i=1}^{I} x_{in}^{(C)},\ t_n^{(E)} = \sum_{i=1}^{I} x_{in}^{(E)},\ t_n^{(T)} = \sum_{i=1}^{I} x_{in}^{(T)},\ n = 1, \ldots, N . \tag{12}$$

By contrast, a conjunctive label for the analogies case would be,

$$\sum_{i=1}^{N} (\theta^{(C)}-\beta_n^{(C)})x_n^{(C)} + \sum_{i=1}^{N} (\theta^{(E)}-\beta_n^{(E)})x_n^{(E)} + \sum_{i=1}^{N} (\theta^{(T)}-\beta_n^{(T)})x_n^{(T)} + \sum_{n=1}^{N} (\theta^{(CE)}-\beta_n^{(CE)})x_n^{(C)}x_n^{(E)} +$$
$$\sum_{n=1}^{N} (\theta^{(CT)}-\beta_n^{(CT)})x_n^{(C)}x_n^{(T)} + \sum_{n=1}^{N} (\theta^{(ET)}-\beta_n^{(ET)})x_n^{(E)}x_n^{(T)} + \sum_{n=1}^{N} (\theta^{(CET)}-\beta_n^{(CET)})x_n^{(C)}x_n^{(E)}x_n^{(T)} , \tag{13}$$

and its corresponding sufficient statistics would be those in (11) and (12), along with

$$s_i^{(CE)} = \sum_{n=1}^{N} x_{in}^{(C)}x_{in}^{(E)} ,\quad s_i^{(CT)} = \sum_{n=1}^{N} x_{in}^{(C)}x_{in}^{(T)} ,$$
$$s_i^{(ET)} = \sum_{n=1}^{N} x_{in}^{(E)}x_{in}^{(T)} ,\quad s_i^{(CET)} = \sum_{n=1}^{N} x_{in}^{(C)}x_{in}^{(E)}x_{in}^{(T)} ,\quad i = 1, \ldots, I , \tag{14}$$

and

$$t_n^{(CE)} = \sum_{i=1}^{I} x_{in}^{(C)}x_{in}^{(E)},\ t_n^{(CT)} = \sum_{i=1}^{I} x_{in}^{(C)}x_{in}^{(T)},$$
$$t_n^{(ET)} = \sum_{i=1}^{I} x_{in}^{(E)}x_{in}^{(T)},\ t_n^{(CET)} = \sum_{i=1}^{I} x_{in}^{(C)}x_{in}^{(E)}x_{in}^{(T)},\ n = 1, \ldots, N . \tag{15}$$

Thus, the conjunctive version would result in 4 more subscale scores per person than the additive version, along with several more item statistics.

A further consequence of the statistical theory behind the *CRF* is that each sufficient statistic reflects its corresponding parameter in a direct and reasonable way. For the analogies case each person's subscale sufficient statistic is positively related to its corresponding parameter's maximum likelihood estimate *(MLE)*. Likewise, each item's sufficient statistic is negatively related to its corresponding parameter's *MLE*. The same direct and reasonable connection between sufficient statistics and parameter estimates holds for all *CRF* models.

Like the Rasch model, the analogies *CRF* model (13) allows for both individual differences and item differences. Simplified versions that focus on only persons or only items may be formed and analyzed simply by setting the appropriate parameters to zero. For example, suppose that two sets of item triplets were to be used and only differences between the two sets were of interest. Individual differences could be excluded completely by excluding all person parameters from (13) and ignoring all person sufficient statistics. This would make the resulting analysis simpler (although it might also reduce power, in analogy to ignoring individual differences in analysis-of-covariance settings).

Conjunctive models for the other two examples may be constructed, interpreted, restricted, and extended as in the analogies case. A conjunctive label for the item chain learning example would have the form,

$$\sum_{m=1}^{M} (\theta^{(G)}-\beta_m)x_m + \sum_{m=1}^{M-1} (\theta^{(D)}-\beta_{m,m+1})x_m x_{m+1} \ . \tag{16}$$

The nonconjunctive label would be the same as that for the Rasch model and would lead to sufficient statistics of the form,

$$s_i^{(G)} = \sum_{m=1}^{M} x_{im} \ , \quad i = 1, \ldots, I \ , \tag{17}$$

and

$$t_m^{(G)} = \sum_{i=1}^{I} x_{im} \ , \quad m = 1, \ldots, M \ , \tag{18}$$

which are the same as these for the Rasch model. Sufficient statistics for the conjunctive version would include those in (17) and (18), along with

$$\sum_{m=1}^{M-1} x_{im}x_{i,m+1}, \ i = 1, \ldots, I \ ; \ \sum_{i=1}^{I} x_{im}x_{i,m+1}, \ m = 1, \ldots, M-1 \ . \tag{19}$$

Figure 3 is a diagram of the CRF model corresponding to (16). Each individual parameter has a causal effect on each of its corresponding conjuncts, as shown. The effects of item parameters on their corresponding conjuncts are also shown—their opposing effects relative to ability parameters are indicated by minus signs. Similar path diagrams could be drawn for the other *CRF* models as well.

Turning finally to models for replicated tests, an additive pretest-posttest model for binary items would have the label,

$$\sum_{k=1}^{K} (\theta^{(pre)}-\beta_k^{(pre)})x_k^{(pre)} + \sum_{k=1}^{K} (\theta^{(post)}-\beta_k^{(post)})x_k^{(post)} \ , \tag{20}$$

and sufficient statistics of the form

$$s_i^{(pre)} = \sum_{k=1}^{K} x_{ik}^{(pre)} \ , s_i^{(post)} = \sum_{k=1}^{K} x_{ik}^{(post)} \ , \quad i = 1, \ldots, I \ ;$$

$$t_k^{(pre)} = \sum_{i=1}^{I} x_{ik}^{(pre)} \ , t_k^{(post)} = \sum_{i=1}^{I} x_{ik}^{(post)} \ , \quad m = 1, \ldots, M \ . \tag{21}$$

The conjunctive version would have a label of the form,

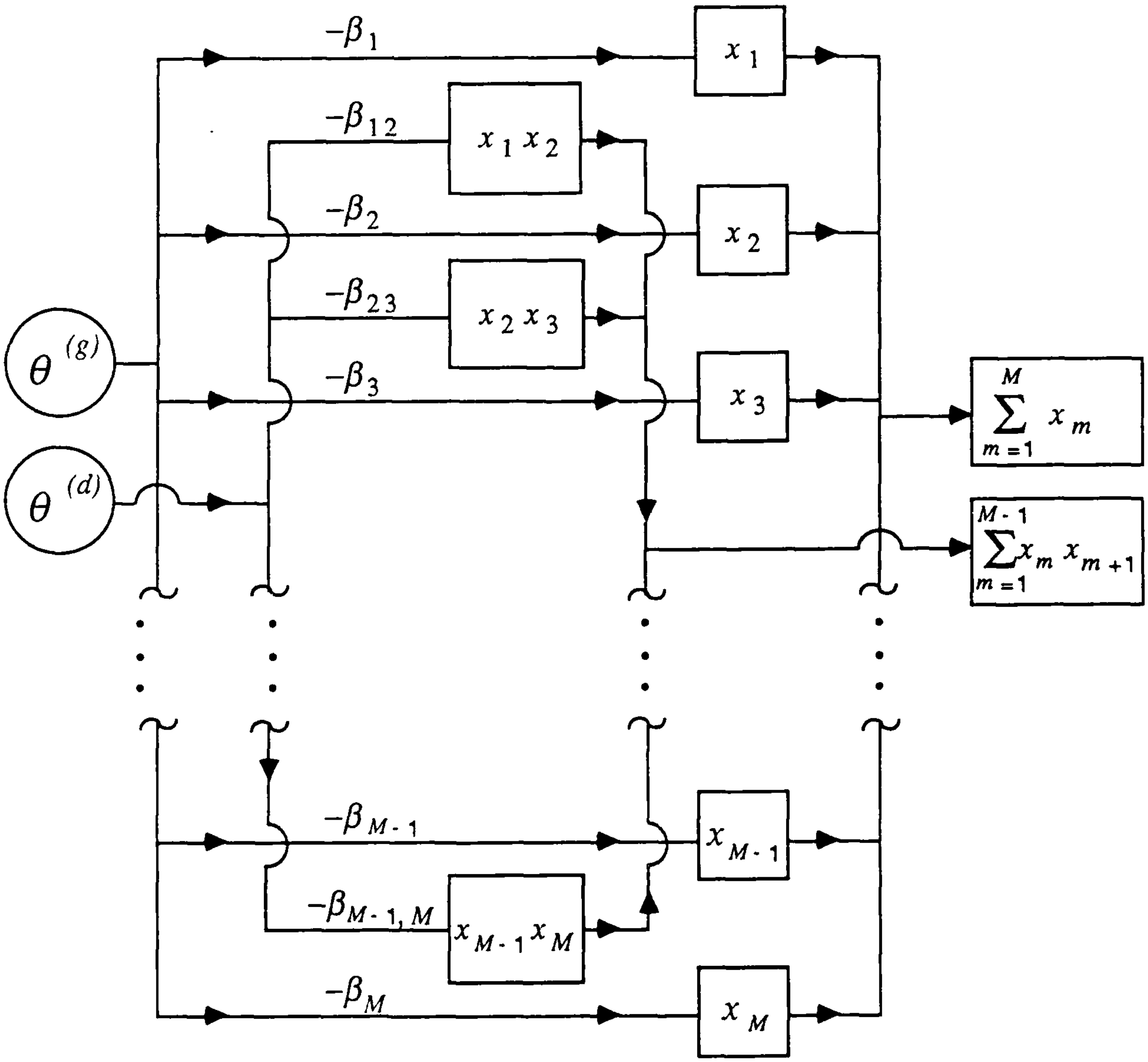
$$\sum_{k=1}^{K} (\theta^{(pre)}-\beta_k^{(pre)})x_k^{(pre)} + \sum_{k=1}^{K} (\theta^{(post)}-\beta_k^{(post)})x_k^{(post)} + \sum_{k=1}^{K} (\theta^{(pre,post)}-\beta_k^{(pre,post)})x_k^{(pre)}x_k^{(post)} \ , \tag{22}$$

and sufficient statistics including the additive statistics in (21), along with conjunctive statistics of the form,

$$s_k^{(pre,post)} = \sum_{k=1}^{K} x_{ik}^{(pre)}x_{ik}^{(post)} \ , \quad i = 1, \ldots, I \ ; \quad t_k^{(pre,post)} = \sum_{i=1}^{I} x_{ik}^{(pre)}x_{ik}^{(post)} \ , \quad k = 1, \ldots, K \ . \tag{23}$$

Labels for continuous model data, such as the pretest-posttest data shown in Figure 2, have a different form than (8). One possible general family is reported elsewhere (Jannarone, 1986), but its relatively

Figure 3. Item Chain Learning Structure



complex form will not be described here.

Procedural guidelines. This section gives some simple guidelines for statistical model assessment. The focus will be on evaluating conjunctive models over and above additive alternatives. Detailed descriptions of efficient procedures are available elsewhere (Jannarone, 1986; 1988a, b, c; Jannarone, Laughlin, & Yu 1988). Computer programs are also available upon request.

Beginning with the item chain learning model, the additive version is simply a Rasch model. The Rasch model, like other *CRF* models leads to relatively simple estimation procedures. The simplest way to evaluate the Rasch model is to directly evaluate associations between external criteria and its item and/or person sufficient statistics. Correlating number-correct scores with external criteria is one such method. More efficient (maximum likelihood or conditional maximum likelihood) estimates than sufficient statistics may be obtained *via* iterative estimation procedures. These procedures are simple (relative to maximum likelihood factor analysis procedures, for example), because the Rasch model belongs to a well-behaved statistical family (see *Sound* versus ad hoc *measurement* below). Efficient Rasch model estimation is especially simple and fast, because its item parameter estimates are obtained independently of each other. Efficient estimates for other *CRF* models are slightly harder to get, because their item parameters are interrelated. The most complicated *CRF* example in this report is the item chain learning model, because all of its items are linked together. The other two models are easier to evaluate, because fewer items are linked together (each pretest-posttest item in the replicated test example and each item triplet in the analogies example).

The easiest conjunctive *versus* additive comparison procedure to describe is for the item chain learning model case. Returning to Table 3, its rows indicate how persons' nonadditive statistics break down samples over and above their additive statistics, as stated earlier. As before, suppose that evaluating associations between the learning test scores and scores on an external variable were of interest. The simplest way to evaluate the conjunctive model over and above the Rasch model would be to test for subgroup differences within each row of Table 3. If the external criterion were binary then tests for equality of binomial proportions (Fleiss, 1981) could be used; if the external criterion were continuous then one-way analyses-of-variances (*ANOVAs*) could be used; multivariate *ANOVAs* could be used for multiple continuous criteria; and so on.

Other more efficient procedures for additive *versus* conjunctive model evaluation are easy to derive, but have not yet been worked out. For example, combining the test for each row of Table 3 into a single global test would be useful. Tests would also be useful for comparing the two models in the absence of external criteria. The simple form of *CRF* models guarantees that (asymptotic likelihood ratio—see Lehmann, 1986) tests could be derived for such cases. Such tests have not yet been worked out, however.

For the analog example, the Total item scores may be regarded as criterion variables and the two subtask scores as predictor variables. Appropriate conjunctive *versus* additive model comparisons may then be based on the corresponding sufficient statistics in (14) and (11). As in the Table 3 case, subgroups would be first obtained by breaking down equal additive sufficient statistics groups according to conjunctive statistic values. For example, among those having scores of 6 and 6 on the two subtask scores, seven subgroups having different cross-product scores could be obtained. Tests for association between these seven subgroups and the Total scores could then be performed (perhaps with a nonparametric version of a one-way *ANOVA* test—see Lehmann, 1975). If the correct model were additive then the seven subgroups would be expected to have the same mean Total scores; otherwise they would not.

Since the analog case involves two additive predictor statistics rather than one, many more individual tests are involved than in the item chain learning case. Thus, a global test that combined the individual tests for group differences would be even more useful. Such a test is currently being developed (Jannarone, 1988c).

Turning next to the pretest-posttest case, if the K measures were binary then additive *versus* conjunctive comparisons would be similar to those for the analog case. For each pretest-posttest sufficient statistic combination, distinct subgroups could be defined by distinct cross-product sufficient statistic values. Conjunctive *versus* additive comparison procedures would focus on assessing external

variable differences for these subgroups. If the K component measures were continuous then pretest-posttest correlation coefficients could first be evaluated for each person. Associations of the coefficients with external criteria, over and above those from the additive pretest and posttest measures, could then be assessed (using a suitable *nonparametric* procedure).

One potential misuse of correlation coefficients should be mentioned. Suppose that pretest-posttest correlations were computed for all persons in a sample; a test for significance of each coefficient were performed; and the number of significant tests greatly exceeded chance levels. Concluding that a conjunctive rather than an additive structure held would *not* be correct in that case. The reason is that differing subtask difficulties could cause the correlations to be high. For example, most individuals could have scatterplots that looked like Figure 2a, simply because the subtasks ranged from easy to hard for all persons. The way to avoid such artifacts is to assess whether *individual differences* in correlations add useful information, which is why such tests have been featured in this section.

I will end this section by discussing the process of aligning cognitive with psychometric models and pointing out its importance. Both fields seem to have developed together, in alternating cycles of increased generality and improved alignment. Thurstone's remarkable success at jointly developing both his psychometric (multiple factor analysis) model and his cognitive (primary mental abilities) model is an excellent example. The *psychometric* side of Thurstone's alignment process included a sublime alignment device: the use of factorial rotation (Thurstone, 1947; Meredith, 1977). His *cognitive* work, which was more subtle and perhaps more important, focused on selecting items that fit his model. The following excerpt from Thurstone's (1938) monograph describes the cognitive side of the process.

> In the exploratory study that we are reporting in this monograph we did not have the advantage of orientation about any known landmarks. Consequently, the tests in the present study were often more complex as to factorial composition than we had anticipated. The tests have been constructed for the subsequent studies as more nearly pure in that some of them could be designed so as to feature one factor with little admixture of others. This process will continue for some time until we shall be able to prepare psychological tests that involve only one or two factors instead of three, four, or five, as is the case with most of the tests in current use.

The most productive development of conjunctive models would entail an alignment process much like the one that Thurstone used. For a given cognitive domain items would first be constructed with a particular conjunctive model in mind. The items would then be empirically evaluated against the model, with well-fitting items being retained and others being discarded. Still other items would be introduced into the process that were similar to those that had been retained, and so on. The *psychometric* part, by contrast, would entail fitting slightly different psychometric models to reflect the slightly different nature of the retained items, and so on.

The potential gains from such an alignment process are much greater than those would be from simply testing conjunctive models against existing additive measures. The reason is that most existing measures have come from a long process of selecting items to fit additive models. It would thus be surprising to find that conjunctive models added much to additive model explanatory power in such cases.

Some Theoretical Issues

Conjunctive versus *compensatory structure*. The term "compensatory" was introduced (Sympson, 1977) to describe cases where an individual's deficiency in one component trait can be overcome by superiority in another. For example, factor analysis and *LISREL* (Jörekog & Sörbom, 1984) models are compensatory in that persons' factor values have additive effects on their component item/subtest scores. As a consequence of additivity, high levels of one factor can compensate for low levels of another. Compensatory and additive models are thus equivalent on the one hand, whereas

conjunctive and nonadditive models are equivalent on the other hand. Compensatory models clearly dominate both early psychometric history and current psychometric practice. Most notably, the classical-test, Spearman, Thurstone, general-linear, Rasch, and logistic models are all compensatory.

The compensatory tradition has been largely unchallenged over the years, perhaps for three main reasons: (1) additive systems explain scientific data well as a rule, and psychometric modeling is no exception; (2) estimation and hypothesis testing procedures based on additive psychometric models are relatively simple; and (3) established additive models tend to reify themselves by encouraging researchers to retain only additive data. For these reasons compensatory models will remain prominent in psychometric modeling, even after compelling alternative models and methods become established.

With all of its virtues, compensatory psychometric modeling carries certain liabilities. First, cognitive psychologists have become increasingly interested in certain *conjunctive* (noncompensatory) tasks (Embretson, 1985; Pellegrino & Glaser, 1979; Pellegrino, Mumaw & Shute, 1985; Sternberg, 1977). Second, all of the examples that I presented earlier are conjunctive rather than compensatory, as are many related instances. Finally, compensatory measurement only reflects how persons perform, not how they *react* to items. This subtle but important distinction will be described next.

Noninvasive versus *reactive measurement*. This section might be subtitled, "An assault on an axiom", the target being test theory's local independence assumption. Given a (possibly multivariate) latent trait value for a person, local independence assumes that all item (or subtest) scores for that person will be mutually independent. Local independence is considered to be test theory's principal axiom (Lazarsfeld, 1958, Lord & Novick, 1968) for several reasons. First, local independence leads to simple mathematics: if items are independent then their joint probabilities are products of their marginal probabilities. Second, local independence brings focus to the test score analysis process. Given that all item dependencies can be explained by person parameters, evaluating person parameters will be sufficient for describing all elements that the items have in common (Lord & Novick, 1968).

The main substantive feature of local independence is its noninvasive measurement property. No matter how a person responds to an item, local independence guarantees that the person will respond to future items in the same way. The item measurement process itself is thus assumed to have no measurable effects on the person's future behavior. (That future item scores will not depend on *whether or not an item was presented* is also assumed insofar as binary item nonresponses are scored as FAILs.) The noninvasive feature implies that the testing process yields independently distributed (*ID*) item scores for each person. The availability of *ID* item scores, in turn, leads to simple and effective procedures for estimating latent traits, evaluating reliability and validity, and so on. In addition, local independence implies that observed individual and treatment differences cannot be caused by the measurement process itself, thus removing a cumbersome confound from the inference process.

Noninvasive measurement has it drawbacks, however, especially in certain developmental settings. Suppose that a person were presented a sequence of information by a tutor in a way that at once evaluated the person, taught the person, and governed how future information was to be presented. Any reasonable explanatory model of such a sequence would necessarily allow the person to change during the process. That is, a reasonable model would allow for local *dependence*. Moreover, by permitting local dependence, such a model would allow persons to *react* to items, thus making them measurably different after taking an item than before. Such might be the case in other repeated measurement settings, as in the replicated test and item chain learning examples that were given earlier.

One way to reflect change and yet preserve local independence is to allow persons' latent traits to change over time. Some of the additive models that I presented earlier allow for such changes. For example, pretest-posttest Rasch and classical models allow for two different traits to be measured at two different time points. Other more detailed models also permit change but preserve local independence (Bieber & Meredith, 1985; Jöreskog & Sörbom, 1977). Such models also can be quite useful in reflecting change (Rogosa & Willett, 1982). For example, a variety of people could have pretest-posttest scatterplots that were similar to the one in Figure 2c, but distinct from each other in terms of pretest, posttest differences. Measuring such differences could be useful, especially if they were related to other interesting variables.

However, merely allowing distinct traits to govern different item responses is not enough in some cases. For example, no such model could reflect the individual differences that appear among the Figure 2 graphs, because all five change scores are the same. The strategies indicated in the other two examples likewise cannot be reflected by merely assigning distinct traits to each item.

The fact that both local dependence and multidimensionality can reflect change has led some authors to conclude that the two are related (Andrich, 1984; Goldstein, 1980; Hambleton, Swaminathan, Cook, Eignor, & Gifford, 1978). Some have speculated that locally dependent models are necessarily multidimensional, although I have been able to construct viable counter examples (Jannarone, 1988a). A more complementary relationship between the two may exist, however, in that local dependence can perhaps always be explained away by introducing additional traits. In the pretest-posttest case, for example, all of the graphs in Figure 2 might be explained by providing for: (a) one distinct trait for each subtask; and (b) other distinct traits that would somehow describe how different persons might have have different strategies, even if they had the same prescore patterns.

Locally dependent and multivariate test models could represent two alternatives, then. The multivariate alternative would provide for predicting each person's performance even in learning settings, provided that the person's latent traits were known. The locally dependent alternative, in turn, could allow for measuring item dependencies that could not be reflected by existing multivariate models.

One major practical problem separates the two, however. Locally independent alternatives to conjunctive measurement are simply not yet available. In addition, the prospects for such models may not be good. For example, no viable multivariate models for reflecting the kinds of strategies in the pretest-posttest example come to mind. Moreover, even if such models could be formulated their resulting statistical procedures might not be sound for certain technical reasons (see below). Therefore, locally dependent measurement procedures may remain useful, at least until viable locally independent alternatives have been worked out.

I will end this argument for considering locally dependent alternatives with a physical measurement analogy. The reaction of persons to items somewhat resembles the reaction of particles to measurement in quantum physics. Physicists have found that certain particle measurements are always invasive. Moreover, such measurements of one property tend to change values of other properties in uncertain ways. It seems like physicists are saying that (a) particles always "notice" when they are being observed and "decide to" react by changing their nature; and (b) modern models are unable to predict how they will react. (A more basic question is whether or not models *could* be formed that would completely describe their reactions—see Suppes, 1976). In terms of the previous discussion, some measurements ***always*** cause particles to react unpredictably, like persons do in learning settings. Most psychologists seem to believe that human dynamics are far more complex than those of atomic particles. It would be ironic, then, if psychologists were to assume that humans ***never*** react to solving test items. Yet, this is precisely the assumption behind test theory's local independence axiom.

I have chosen the term ***reactive*** rather than ***interactive***, because locally dependent reactions are distinct from item-by-person *ANOVA* interactions. One distinction is only semantic: strictly speaking, when an item is taken no interaction is possible—the person can react to the item but not the item to the person. The second distinction is between the role assigned to interactions in statistics and the role of locally dependent measures. Typical *ANOVA* interaction formulations treat underlying data as both replicable and *ID*. (This also appears to be the case with interactive item response models, *e.g.* Spada and McGaw, 1975.) By sharp contrast, locally dependent models cannot permit replicable, *ID* observations. Not only can the person change while taking an item, but item scores can be statistically dependent as well. (The interaction, reaction distinction also seems related to the frequentist *versus* Bayesian debate in statistics (Savage, 1954; Neyman, 1977)—for example making frequentist inferences about a person being measured reactively would seem to be much more awkward than making Bayesian inferences.)

So far several terms have been used to describe models that are equivalent. It may be useful to group them together at this point for clarity. With minor technical exceptions, additive, traditional, compensatory, noninvasive and locally independent models represent one alternative, whereas nonadditive, conjunctive, reactive and locally dependent models represent the other. Except where noted,

such terms will be used interchangeably in the sequel.

Sound versus ad hoc *measurement.* So far I have only contrasted conjunctive with compensatory ability measurement. However, conjunctive as well as compensatory versions of several different models are possible, including factor analysis, logistic, Rasch, and binomial models. As a result the prospects for conjunctive versions of these different models are worth noting. Although I have introduced conjunctive versions for all four models (Jannarone, 1986), I have focused on developing only conjunctive Rasch versions so far. In this section I will indicate the main advantage as well as some disadvantages of conjunctive Rasch extensions, relative to other possibilities.

Several bases could be considered for evaluating competing models, but I will focus on two: axiomatic soundness or *substantive validity* and statistical soundness or *procedural viability*. From an axiomatic viewpoint a mathematical model can range from being very specific to very general. Very specific models are also called strong models (Lord & Novick, 1968), because they make restrictive and falsifiable assumptions about nature. Conversely, more general models are called weak, but they can reflect a broader array of natural events. Ideally, a family of test models would be available with members ranging from very specific to very general. Procedures for selecting the best family members for particular situations would also ideally be available. The family of test models does not line up according to such a neat generality ordering, however. Instead, the family tree branches off in a few axiomatic directions, each having the following relative strengths and weaknesses. (The references in the next four paragraphs are primary—more refined descriptions appear in Gulliksen, 1950, Lord & Novick, 1968, and Thissen & Steinberg, 1986.)

Beginning with test models for continuous component scores, the oldest and strongest is the classical test model (Spearman, 1904). The classical model's axioms provide for independently and identically distributed *(IID)* as well as unidimensional item scores. Spearman's (1904, 1927) factor analysis model is more general than the classical model, because it relaxes the assumption of equal item weighting. Spearman's model also allows component measures to have unequal difficulties, because it is based on standardized rather than raw component measures. Thurstone's (1932) factor analysis model is still more general than Spearman's, because it relaxes the unidimensionality assumption. Recent conjunctive extensions of multivariate normality (Jannarone, 1986) point toward conjunctive versions of all three continuous test models. No continuous conjunctive versions have yet been refined, however.

Applications of continuous models to binary data have historically resulted in serious violations of the *IID* error assumption. Item response theory for binary items has been developed as a result. Item response (*IR*) models make up a distinct branch of test theory, because they reflect binary rather than continuous responses and they have their own generality ordering.

The strongest *IR* model, called the binomial model (Keats & Lord, 1962), requires that all items have the same characteristics, much like the classical test model. The Rasch model (Rasch, 1980) allows items to have different difficulties, and still weaker *IR* models allow items to vary in discriminating power as well. (An item's discriminating power evaluates its change-in-difficulty to change-in-ability ratio.) One of these is called the normal ogive model (Ferguson, 1942; Lawley, 1943), and the other is called the two-parameter logistic model (Birnbaum, 1958). The Rasch and binomial models are like the classical model in that they yield ability estimates that are unweighted sums of item scores. By contrast, the two-parameter logistic model yields weighted sums of item scores as ability estimates, like Spearman's model for continuous measures. Still weaker *IR* models allow different items to have different wild guessing probabilities. The most popular among these is the three-parameter logistic model (Birnbaum, 1968).

Multidimensional versions of Rasch models (Fischer, 1973; Whitely, 1980) and logistic models (McKinley & Reckase, 1983) have also been introduced. Categorical versions of *IR* models have appeared as well (Andrich, 1978; Bock, 1972; Samejima, 1969). Finally, conjunctive versions of binomial, Rasch, logistic, and multidimensional IR models have all been introduced (Andrich, 1985; Fischer & Formann, 1982; Embretson, 1984; Jannarone, 1986a, 1988a; Kempf, 1977; Lord, 1984; Spray & Ackerman, 1986). However only the conjunctive Rasch versions have been successfully developed Embretson, 1984; Jannarone, 1988a).

From an axiomatic viewpoint, it would be best to develop conjunctive versions of the most general models available. The most general versions could reflect the broadest array of test responses, and they could also identify simpler versions as special cases. Thus, multivariate logistic models and factor analysis models would be the soundest choices for conjunctive development, on purely axiomatic/substantive validity grounds.

From the viewpoint of statistical/procedural viability, however, both factor analysis and logistic models are not as sound. Statistical problems may exist for factor analysis and logistic models, because they do not belong to a very sound and broad family of statistical models called the *multiparameter exponential family* (Anderson, 1980; Lehmann, 1983). Exponential family likelihoods are easy to use because their exponents are additive functions of parameters. Exponential family members have maximum likelihood estimates that are unique; iterative estimation procedures that always converge; and hypotheses test statistics that have known asymptotic distributions (Andersen, 1980; Lehmann, 1983). Useful Bayes estimates (Jannarone, Laughlin, & Yu, 1988) are also easy to derive for exponential family models. Multiparameter exponential family members also have some useful conditional probability properties, including: (a) easy provisions for statistical control—the effects of nuisance parameters can be removed simply by conditioning on the nuisance variables' sufficient statistics; and (b) and provisions for conditional maximum likelihood (*CML*) estimation—*CML* estimates can sometimes be obtained much more quickly and simply than maximum likelihood estimates.

The Rasch model's exponential family membership has led to simple estimation procedures, relative to those for weaker models. For example, no local maximum, nonconvergence, or nonidentifiability problems have occurred with Rasch models. The Rasch model has another feature—stemming from the exponential family's statistical control properties—that is very attractive for testing applications. Item parameters and individual parameters for the Rasch model can be estimated completely separately from each other (Rasch, 1980). The conjunctive Rasch models that I have developed (Jannarone, 1986, 1988a) are not only exponential model members but they share the same sound estimation properties as well.

By contrast, nonexponential family members have potentially serious estimation problems. For example, factor analysis/*LISREL* models are not exponential family members, because their likelihoods involve cross-products among parameters. Such models are known to have identifiability, local solution, and nonconvergence properties (Long, 1983). Logistic item response models are also potentially problematic, because they too have products of parameters in their likelihoods. Thus, no guarantee exists that logistic parameter estimates are indeed optimal or even unique.

(I should point out that equating soundness with exponential family membership is a bit simplistic. For example, some nonexponential family members such as the principal components model have very sound least squares properties (Eckard & Young, 1936). Also, exponential model estimates are not perfect—they tend to be biased in ways that persist even in large samples, for example (Anderson, 1982). In addition, existing factor analysis and logistic model procedures must be sound for the most part, or else they would have long since been abandoned. However, debugging complex test model estimation algorithms is much easier when they carry an unconditional guarantee of correct convergence—otherwise it is hard to identify when model problems begin and programming mistakes end.)

In sum, I have described two criteria for evaluating the soundness of test models: substantive validity and procedural viability. In terms of substantive validity, multivariate factor analysis and logistic models would be ideal for conjunctive model development, because they are the most general. In terms of procedural viability, however, Rasch type models are much preferred. So far I have decided to develop Rasch conjunctive models because of their sound statistical properties. Other conjunctive approaches may also prove to be reasonable, however.

I will now briefly mention some related models. Conjunctive models have been formulated by Andrich (1978; 1985), Bryk and Raudenbush (1987), Embretson (1984), Fischer & Formann (1982—a Rorshcash model with 'technical items') Jannarone & Roberts (1984), Jöreskog (1984—*LISREL* with correlated errors) Kempf (1977), Lord (1984), McDonald (1967—nonlinear factor analysis with latent trait cross-products) Rogosa & Willett (1982—change scores having correlated errors) and Spray & Ackerman (1986). Embretson's model appears to be both sound and completely worked out—it is

indeed a Rasch conjunctive model, but without conjunctive individual difference parameters. Andrich's and Spray & Ackerman's test models, along with Bryk & Raudenbush's and Rogosa & Willett's growth models appear to be sound, but their locally dependent measurement procedures have not yet been developed. Jannarone & Roberts' (1984) method is unsound for reasons that are related to the misuse of continuous models for binary data. (For reviews of this and related *configural scoring* methods, see Jannarone & Roberts, 1984, and Jannarone, 1986). The remaining models may lead to procedural problems because they are either improperly specified or they are not members of the exponential family.

Additive versus *nonadditive measurement.* I have suggested previously that test models are nonadditive whenever they are locally independent, provided that they belong in the conjunctive Rasch family. In this section I will describe the nonadditivity/conjunctivity connection more precisely. I will also describe how additivity can severely constrain item response function form.

Given membership in the conjunctive Rasch family, if a pair-wise cross-product appears in a model's *CRF* label (8), then the items will be pair-wise locally dependent. The converse is not true, however: a pair-wise cross-product may appear in (8) with the items being pair-wise locally independent. The reason is related to the fact that three binary variables can be pair-wise independent yet mutually dependent. Table 4 gives one such example of a model without individual differences. It turns out that the *CRF* label corresponding to Table 4 includes all pair-wise item cross-products, each having nonzero coefficients. Yet, x_1 and x_2 in Table 4 are clearly independent. The relationship between pair-wise nonadditivity and local independence is clearer if *CRF's* are restricted to include only first-order and second-order conjuncts. In that case nonadditivity both is necessary and sufficient for local independence.

An analogous relationship between cross-products and local independence holds in the continuous case. Multivariate normal models restrict sufficient statistics to include only additive and second-order terms in the observed scores. Given a multivariate normal model, then (such as the bivariate normal pretest-posttest model that was suggested for Figure 2), pair-wise local independence is equivalent to the exclusion of item cross-products from parameter sufficient statistics (*i.e.* zero-valued correlation coefficients.) In the absence of multivariate normality, however, the relationship between cross-products and local independence is more complex, as in the binary case.

Turning next to item response function (*IRF*) form, *IRFs* such as those in Figure 1 specify the relationships between item passing probabilities and ability levels. All of the traditional additive item response models lead to item response functions that are strictly increasing. Increasing *IRFs* are too restrictive to describe two behaviors that I mentioned earlier: (a) cases where clearly brighter students perform worse than less bright students, because they "think themselves into a jam" (b) cases where students/trainees perform worse after some training than they did at the outset. Only nonadditive models that permit nonmonotone *IRFs* have the needed flexibility.

On-line versus *off-line measurement.* This section describes some prospects for estimating parameters nearly instantly. Such prospects point toward developing both conjunctive and additive models in settings that require *on-line* rather than *off-line* measurement. For example traditional educational testing allows tests to be taken at one point but abilities and item difficulties to be estimated off-line at some later point. By contrast tailored testing or computer aided instruction would require parameter estimates to be updated each time a person reacted to an item. Neural and machine learning models must also allow for fast parameter updating (*i.e.* internal representation updating) in order to be practical (Jannarone, Yu, & Takefuji, 1988). On-line test parameter estimation prospects thus point toward a much broader substantive base for test models than merely traditional testing.

In their present form procedures for estimating parameters based on the *CRF*, including the Rasch model, have limited potential because they are iterative. Using such procedures for interactive modeling is not practical, because they may take many seconds to converge. Also, the possibility that humans use such iterative procedures to update their learning states is simply out of the question (given that neurons take about 10^5 times longer to function than computer processing units). Thus, iterative procedures are limited as either vehicles for real time measurement or models of human

Table 4. An Example of Three Mutually Dependent Yet Pairwise Independent Binary Random Variables.*

		$x_2 = 0$	$x_2 = 1$	
x_1	0	.25	.75	.50
	1	.75	.25	.50
		.50	.50	

* Entries are joint and marginal probabilities that X_3 is 1.

thought.

Far better possibilities for fast estimation new exist, because of some recent developments in computer technology and statistical parameter estimation. These will only be mentioned here—for details see Jannarone, Yu, & Takefuji (1988), Takefuji & Jannarone (1988), and Yu & Jannarone (1988). Briefly, *CML* estimation allows useful estimates for a parameter to be expressed as functions of only three variables: the sufficient statistic for that parameter, along with the lowest and highest values that the statistic can take given the other sufficient statistics' values. Also, for members of the Rasch conjunctive family those three variables may be rescaled so as to always fall between 0 and 1. Consider a 100 x 100 x 100 array representing a million equally spaced points on a cube, having boundaries at 0 and 1 along all three dimensions. Current very large scale integration (*VLSI*) technology allows for such an array to exist on a single chip in the form of a read-only-memory. Moreover, each address in the array could be rapidly accessed (in about 100 nanoseconds). Now consider such a chip with each of its elements containing the known *CML* estimate corresponding to its three independent statistic values. Given such availability on-line estimation would be feasible, since after each person's (or learning machine's) item response sufficient statistics could be quickly updated and their corresponding updated *CML* parameter estimates could be quickly accessed. Prototypes of massively parallel computing modules that implement such estimation procedures in about one microsecond are currently being fabricated (Takefuji & Jannarone, 1988).

In sum, recent developments in statistical estimation theory and *VLSI* technology are pointing toward on-line *versus* off-line measurement capabilities for Rasch conjunctive models. Given such capabilities some new media for conjunctive as well as additive test models—including tailored testing, computer aided instruction, and neural/machine learning—may become feasible.

Conclusion

Future directions. Beginning with some necessary psychometric work, the need for several added statistical procedures has been indicated earlier. Besides that need, developing conjunctive models for categorical rather than strictly binary items seems necessary. Multiple-category extensions of conjunctive models would be useful for at least three reasons. First, scoring multiple choice items as only correct or incorrect can lead to distorted results due to wild guessing. Second, a good deal of useful information may be obtainable from multiple category items. Indeed, much more cognitively interesting multiple choice formats than PASS-FAIL could be considered if more general categorical item response models were available. Such formats could become useful in the analysis of choice and attitude structures as well. The prospects for categorical extensions seem promising (Andrich, 1985; Laughlin & Jannarone, 1986), but some procedural details still need to be worked out.

At a more foundational level, reexamining Luce's (1959) choice axiom in light of the previous *Noninvasive* versus *reactive measurement* discussion might be useful. It seems that the choice axiom could be described in terms of whether or nor current choices depend on previous choices. With that connection to the previous discussion in mind, perhaps categorical extensions of conjunctive models could lead to useful extensions of Luce's logistic choice model as well.

A third prospect is the potential for evaluating conjunctive functions of response speed and response accuracy measures. It seems clear (*e.g.* Bloxom, 1985) that since response latencies can be easily measured in computer aided testing settings, procedures based on such measures should also be developed. I would add that since many parametric models based on latencies belong in the exponential family, viable models based on latency/correctness conjunctions may easily be worked out by using conjunctive models. For example, scoring an item differently if it was answered correctly *and* quickly rather than correctly *and* slowly could be useful. Also, focusing only on latencies in experimental studies rather than including accuracy measures as well has been rightly criticized (Whitely & Barnes, 1979). Simple conjunctive approaches may be useful in such experimental settings as well.

A fourth prospect involves tailored item selection. When the choice of items to be administered depends on previous item performance, item scores will necessarily be locally dependent. For this reason it seems not only natural but essential to model local dependence into tailored testing.

Finally, the necessary framework for aligning of psychometric models to conjunctive settings seems to now be available. However, work toward aligning conjunctive *cognitive* substance with conjunctive *psychometric* models has only begun. Major efforts toward identifying suitable settings, screening suitable items, and modifying models as necessary, will be required before conjunctive ability measurement can become useful.

Summary. First, a variety of examples have been used to show how some cognitive traits can be measured conjunctively. These include persons' abilities to (a) combine component skills that may be individually necessary for solving a composite task; (b) learn information at one point and successfully apply it at a later point; (c) positively transfer learned information from one setting to another; and (d) improve knowledge in different ways, depending on initial performance.

Second, a variety of theoretical issues concerning cognitive measurement have been described. These include the distinction between compensatory, additive, locally independent, and noninvasive measurement on the one hand; and conjunctive, nonadditive, locally dependent, and reactive measurement on the other hand. In addition, conjunctive measurement has been contrasted with compensatory measurement on both axiomatic validity and procedural viability grounds. Some issues regarding estimation speed and related prospects have been mentioned as well.

Finally, some future conjunctive directions for both psychometric and cognitive research have been outlined. Among these, the need for related categorical, choice, latency, and tailored testing developments have been mentioned. Above all, the need for coordinated psychometric and cognitive efforts has been stressed.

References

Andersen, E. B. (1980). *Discrete statistical models with social science applications.* Amsterdam: North Holland.

Andersen, E. B. (1982). Latent trait models and ability parameter estimation. *Applied Psychological Measurement, 6,* 445-461.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Andrich, D. (1984, April). The attenuation paradox of traditionalist test theory as a breakdown of local independence in person item response theory. Paper presented at the National Conference of Measurement in Education, New Orleans.

Andrich, D. (1985). A latent trait model for items with response dependencies: implications for test construction and analysis. In Embretson, S. E.(Ed), *Test Design: development in Psychology and Psychometrics.* Orlando, Florida: Academic Press, pp. 245-275.

Bieber, S. L. & Meredith, W. (1986). Transformation to achieve a longitudinally stationary data matrix. *Psychometrika, 51,* 535-547.

Bloxom, B. (1985). Considerations in psychometric modeling of response time. *Psychometrika, 50,* 383-398.

Birnbaum, A. (1958). Statistical theory of tests of a mental ability. *Annals of Mathematical Statistics, 29,* 1285 (abstract).

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In lord, F. M. & Novick, M. R. (Eds.), *Statistical theories of mental test scores.* Reading, Massachusetts: Addison-Wesley.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more numerical categories. *Psychometrika, 37,* 29-51.

Bryk, A. S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101,* 147-158.

Eckard, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 1*, 211-218.

Embretson (Whitely), S. (1984). A general latent trait model for response processes. *Psychometrika, 49*, 175-186.

Embretson, S. E. (1985). Multicomponent latent trait models for test design. In Embretson, S. E.(Ed). *Test Design: Developments in Psychology and Psychometrics.* Orlando Florida: Academic Press, pp. 195-218.

Ferguson, G. A. (1942). Item selection by the constant process. *Psychometrika, 7*, 19-29.

Fischer, G. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*, 359-374.

Fischer, G., & Formann, A. K. (1982). Some applications of logistic latent trait models with linear constraints on the parameters. *Applied Psychological Measurement, 6*, 397-416.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions (2nd Edn.).* New York: Wiley.

Goldstein, H. (1980). Dimensionality, bias, independence and measurement scale problems in latent trait test score models. *British Journal of Mathematical and Statistical Psychology, 33*, 234-246.

Guilford, J. P. (1967). *The nature of human intelligence.* New York: McGraw Hill.

Gulliksen, H. (1950). *Theory of Mental Tests.* New York: Wiley.

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eignor, D. E., & Gifford, J. A. (1978). Developments in latent trait theory: Models, technical issues, and applications. *Review of Educational Research, 48*, 467-510.

Jannarone, R. J. & Roberts, J. S. (1984). Reflecting interactions among personality items: Meehl's paradox revisited. *Journal of Personality and Social Psychology, 47*, 621-628.

Jannarone, R. J. (1986). Conjunctive item response theory kernels. *Psychometrika, 51*, 357-373.

Jannarone, R. J., Laughlin, J. E., & Yu, K. F. (1988). Easy Bayes estimates for Rasch-type models. *Psychometrika* (in review).

Jannarone, R. J. (1988a). Locally dependent models for reflecting learning abilities. *Psychometrika* (in review).

Jannarone, R. J. (1988b). Locally dependent models for Embretson analogy items. *Psychometrika* (in preparation).

Jannarone, R. J. (1988c). Locally dependent models for replicated tests. *Psychometrika* (in preparation).

Jöreskog, K. & Sörbom, D. (1984). *LISREL VI Users' Guide.* Chicago: International Educational Resources.

Jöreskog, K. G. & Sörbom, K. (1977). Statistical models and methods for analysis of longitudinal data. In D. V. Aigner & A. S. Goldberger (Eds.), *Latent variables in sociometric models.* Amsterdam: North Holland.

Jöreskog, K. G. (1978). Statistical analysis of covariance and correlation matrices. *Psychometrika, 43*, 443-477.

Keats, J. A. & Lord, F. M. (1962). A theoretical distribution for mental test scores. *Psychometrika, 27*, 59-72.

Kempf, W. F. (1977). A dynamic test model and its use in the microevaluation of instructional material. In H. Spada & W. F. Kempf (Eds.), *Structural models for thinking and learning.* Vienna: Hans Huber.

Laughlin, J. E. & Jannarone, R. J. (1985). Latent trait items in choice settings. Paper presented at the annual Psychometric Society Meetings, Nashville, Tennessee.

Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal society of Edinburgh, 61*, 273-287.

Lazarsfeld, P. F. (1958). Latent structure analysis. In S. Koch (Ed.) *Psychology: a study of a science, Vol. III.* New York: McGraw-Hill.

Lehmann, E. L. (1975). *Nonparametrics: statistical methods based on ranks*. San Francisco: Holden-Day.

Lehmann, E. L. (1983). *Theory of point estimation.* New York: Wiley.

Lehmann, E. L. (1986). *Testing statistical hypotheses (2nd Edn.).* New York: Wiley.

Long, J. S. (1983). *Covariance structure models: an introduction to LISREL.* Beverly Hills, California: Sage.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, Massachusetts: Addison-Wesley.

Lord, F. M. (1984). Conjunctive and disjunctive item response functions. Educational Testing Service Technical Report No. 84-45-ONR.

Luce, R. P. (1959). *Individual choice behavior*. New York: Wiley.

McCormick, C. & Miller, G. E. (1986). A comparison of mnemonic approaches to learning English vocabulary. Unpublished manuscript.

McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometric Monograph No 15.*

Mckinley, R. L., & Reckase, M. D. (1983). An extension of the two-parameter logistic model to the multidimensional latent space (Research Report Number R83-2). Iowa City: American College Testing Program.

Meredith, W. (1977). On weighted procrustes and hyperplane fitting in factor analytic rotation. *Psychometrika, 42,* 491-522.

Meredith, W. & Tisak, J. (1982). Canonical analysis of longitudinal and repeated measures data with stationary regression weights. *Psychometrika, 47,* 47-67.

Pellegrino, J. W. & Glaser, R. (1979). Cognitive components and correlates in the analysis of individual differences. *Intelligence, 3,* 187-214.

Pellegrino, J. W., Mumaw, R. J., & Shute, V. J. (1985). Analysis of spatial aptitude and expertise. In Embretson, S. E.(Ed), *Test Design: Developments in Psychology and Psychometrics.* Orlando, Florida: Academic Press, pp. 45-76.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press.

Rogosa, D. R. & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement, 20,* 335-343.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph No. 17.*

Spada, H. and McGaw, B. The assessment of learning effects with linear logistic test models. In Embretson, S. E.(Ed), *Test Design: Developments in Psychology and Psychometrics.* Orlando, Florida: Academic Press, pp. 169-194.

Spray, J. A. & Ackerman, T. A. (1986, June). The effects of item response dependency on trait or ability dimensionality. Paper presented at the annual Psychometric Society Meetings, Toronto.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15,* 72-101.

Spearman, C. (1923). *The nature of intelligence and the principles of cognition.* London: Macmillan.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Sternberg, R. S. (1977). *Intelligence, information processing and analogical reasoning: the componential analysis of human abilities.* Hillsdale, New Jersey: Erlbaum.

Suppes, P. (Ed.) (1976). *Logic and probability in quantum mechanics*. Dordrecht, Holland: Reidel.

Sympson, J. B. (1977, July). A model for testing with multidimensional items. Paper presented at the Adaptive Testing Conference, Minneapolis.

Thissen, D. & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51,* 567-577.

Thurstone, L. L. (1932). *The theory of multiple factors.* Ann Arbor, Michigan: Edwards Brothers.

Thurstone, L. L. (1938). *Primary Mental Abilities.* Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple factor analysis: a development and expansion of the vectors of mind*. Chicago: University of Chicago Press.

Whitely, S. E. & Barnes, G. M. (1979). The implications of processing event sequences for theories of analogical reasoning. *Memory and Cognition,* 7, 323-331.

Whitely, S. E. (1980). Multi-component latent trait models for ability tests. *Psychometrika, 45,* 479-494.

Dr. Terry Ackerman
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

Dr. Robert Ahlers
Code N711
Human Factors Laboratory
Naval Training Systems Center
Orlando, FL 32813

Dr. James Algina
University of Florida
Gainesville, FL 32605

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
ISRAEL

Dr. Arthur S. Blaiwes
Code N711
Naval Training Systems Center
Orlando, FL 32813

Dr. Bruce Bloxom
Defense Manpower Data Center
550 Camino El Estero,
Suite 200
Monterey, CA 93943-3231

Dr. R. Darrell Bock
University of Chicago
NORC
6030 South Ellis
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code N-095R
Naval Training Systems Center
Orlando, FL 32813

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. Lyle D. Broemeling
ONR Code 1111SP
800 North Quincy Street
Arlington, VA 22217

Mr. James W. Carey
Commandant (G-PTE)
U.S. Coast Guard
2100 Second Street, S.W.
Washington, DC 20593

Dr. James Carlson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

Dr. Robert Carroll
OP 0187
Washington, DC 20370

Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Hans Crombag
University of Leyden
Education Research Center
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

Dr. Timothy Davey
Educational Testing Service
Princeton, NJ 08541

Dr. C. M. Dayton
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Building
University of Maryland
College Park, MD 20742

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1A226
Piscataway, NJ 08854

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. John M. Eddins
University of Illinois
252 Engineering Research
Laboratory
103 South Mathews Street
Urbana, IL 61801

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

Dr. George Englehard, Jr.
Division of Educational Studies
Emory University
201 Fishburne Bldg.
Atlanta, GA 30322

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

Dr. Pat Federico
Code 511
NPRDC
San Diego, CA 92152-6800

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52240

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rm 529W
1601 W. Taylor Street
Chicago, IL 60612

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Dipl. Pad. Michael W. Habon
Universitat Dusseldorf
Erziehungswissenschaftliches
Universitatsstr. 1
D-4000 Dusseldorf 1
WEST GERMANY

Dr. Ronald K. Hambleton
Prof. of Education & Psychology
University of Massachusetts
at Amherst
Hills House
Amherst, MA 01003

Dr. Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Grant Henning
Senior Research Scientist
Division of Measurement
Research and Services
Educational Testing Service
Princeton, NJ 08541

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Paul W. Holland
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 90010

Mr. Dick Hoshaw
OP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
Department of Psychology
University of South Carolina
Columbia, SC 29208

Dr. Dennis E. Jennings
Department of Statistics
University of Illinois
1409 West Green Street
Urbana, IL 61801

Dr. Douglas H. Jones
Thatcher Jones Associates
P.O. Box 6640
10 Trafalgar Court
Lawrenceville, NJ 08648

Dr. Milton S. Katz
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Prof. John A. Keats
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
University of Texas-Austin
Measurement and Evaluation
Center
Austin, TX 78703

Dr. James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana, IL 61801

Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Daryll Lang
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Jerry Lehnus
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

END
DATE
FILMED
7-88
DTIC